HR data analysis and contract termination prediction

Gabriel Scozzarro

1.0 Introduction

2.0 Preparation and reading data

Data was imported from the set contained inside the file HRDataset_v14.csv

Table 1: Data summary

Name	hr_data
Number of rows	311
Number of columns	36
Column type frequency:	
character	18
numeric	18
Group variables	None

Variable type: character

| skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ï..Employee_Name	0	1	8	25	0	311	0
Position	0	1	3	28	0	32	0
State	0	1	2	2	0	28	0
DOB	0	1	8	8	0	307	0
Sex	0	1	1	2	0	2	0
MaritalDesc	0	1	6	9	0	5	0
CitizenDesc	0	1	10	19	0	3	0
HispanicLatino	0	1	2	3	0	4	0
RaceDesc	0	1	5	32	0	6	0
DateofHire	0	1	8	10	0	101	0
DateofTermination	0	1	0	10	207	97	0
TermReason	0	1	5	32	0	18	0
EmploymentStatus	0	1	6	22	0	3	0
Department	0	1	5	20	0	6	0
ManagerName	0	1	8	18	0	21	0
RecruitmentSource	0	1	5	23	0	9	0
PerformanceScore	0	1	3	17	0	4	0
LastPerformanceReview_Date	0	1	8	10	0	137	0

Variable type: numeric

| skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
EmpID	0	1.00	10156.00	89.92	10001.00	10078.50	10156.00	10233.5
MarriedID	0	1.00	0.40	0.49	0.00	0.00	0.00	1.0
MaritalStatusID	0	1.00	0.81	0.94	0.00	0.00	1.00	1.0
GenderID	0	1.00	0.43	0.50	0.00	0.00	0.00	1.0
EmpStatusID	0	1.00	2.39	1.79	1.00	1.00	1.00	5.0
DeptID	0	1.00	4.61	1.08	1.00	5.00	5.00	5.0
PerfScoreID	0	1.00	2.98	0.59	1.00	3.00	3.00	3.0
FromDiversityJobFairID	0	1.00	0.09	0.29	0.00	0.00	0.00	0.0
Salary	0	1.00	69020.68	25156.64	45046.00	55501.50	62810.00	72036.0
Termd	0	1.00	0.33	0.47	0.00	0.00	0.00	1.0
PositionID	0	1.00	16.85	6.22	1.00	18.00	19.00	20.0
Zip	0	1.00	6555.48	16908.40	1013.00	1901.50	2132.00	2355.0
ManagerID	8	0.97	14.57	8.08	1.00	10.00	15.00	19.0
EngagementSurvey	0	1.00	4.11	0.79	1.12	3.69	4.28	4.7
EmpSatisfaction	0	1.00	3.89	0.91	1.00	3.00	4.00	5.0
SpecialProjectsCount	0	1.00	1.22	2.35	0.00	0.00	0.00	0.0
DaysLateLast30	0	1.00	0.41	1.29	0.00	0.00	0.00	0.0
Absences	0	1.00	10.24	5.85	1.00	5.00	10.00	15.0

3.0 EDA

3.1 Gender analysis

We start analyzing the employees presence by age and gender. The women are more present by 41 units. Women an men has the same average age is which is 42
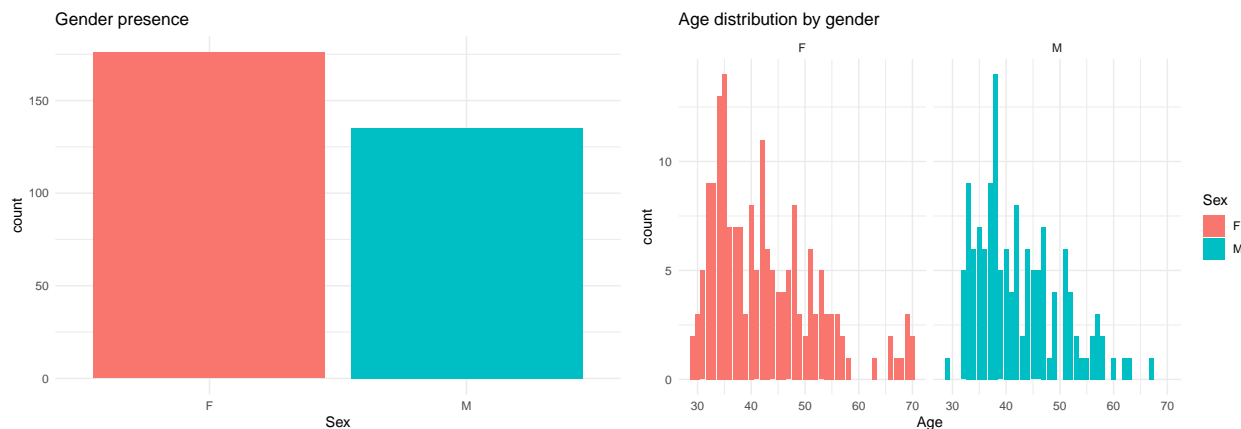
The gender presence by role is as follows:

## Gender presence by role up to date
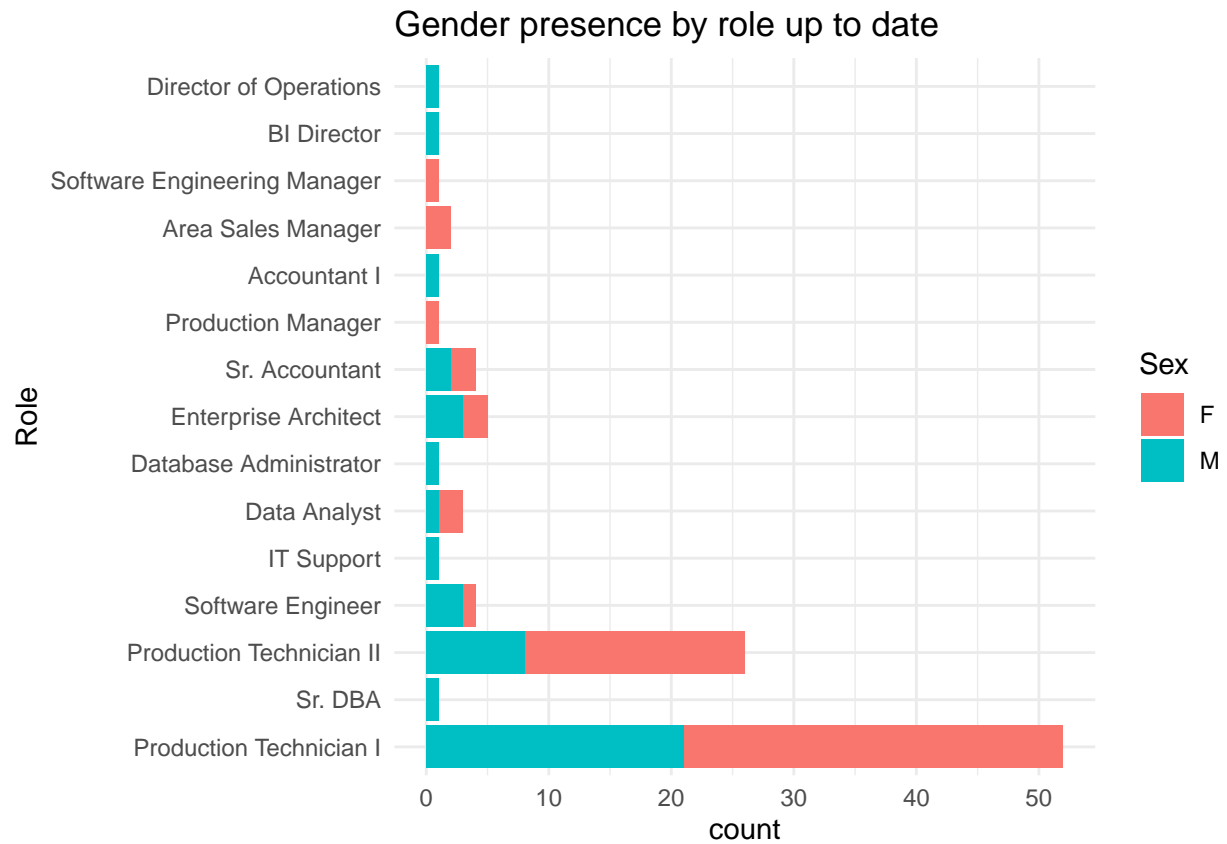
Surprisingly there is a total presence of woman as production manager and sales manager instead men are almost totally prevalent in IT based department.

The enrollment trend by gender from 2008 to 2016 was very much in favor of women. The enrollment since the incorporation was as follows:
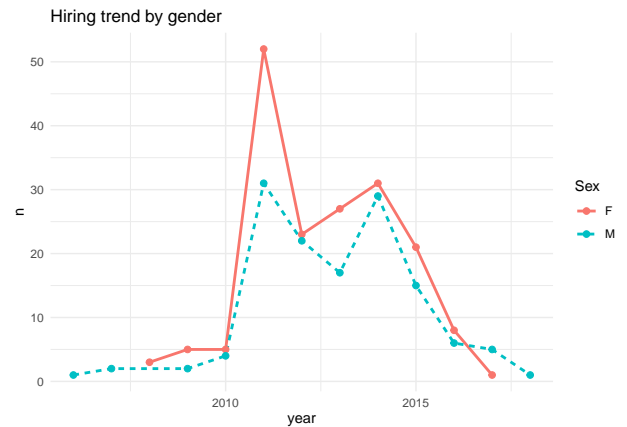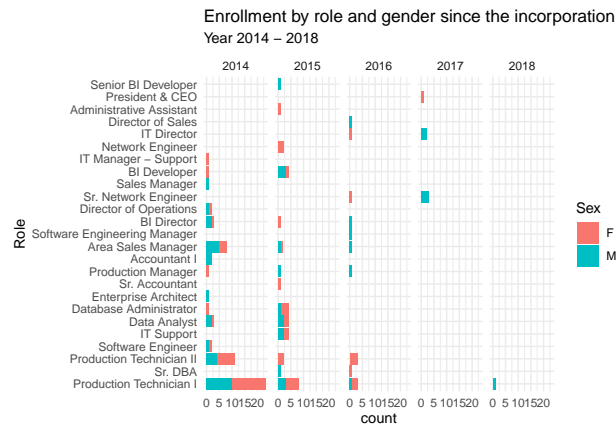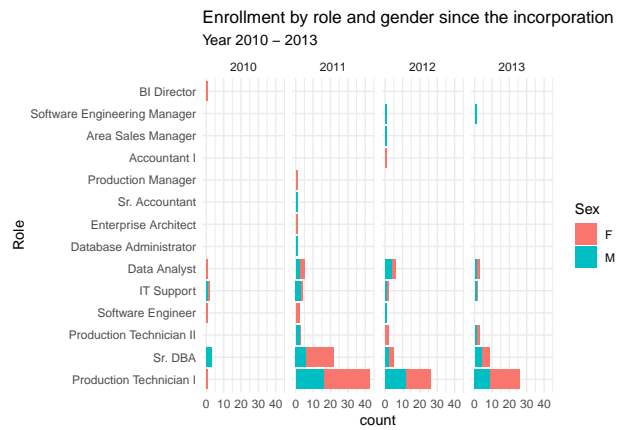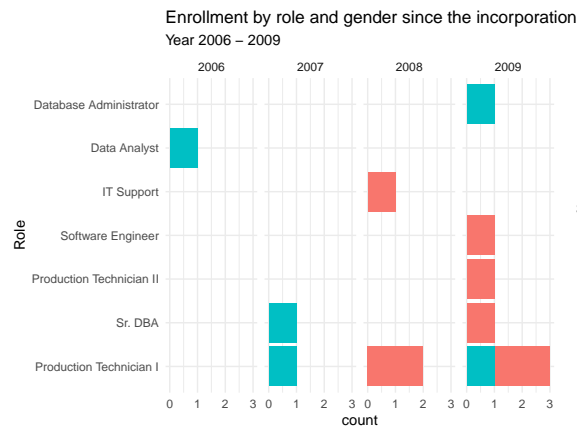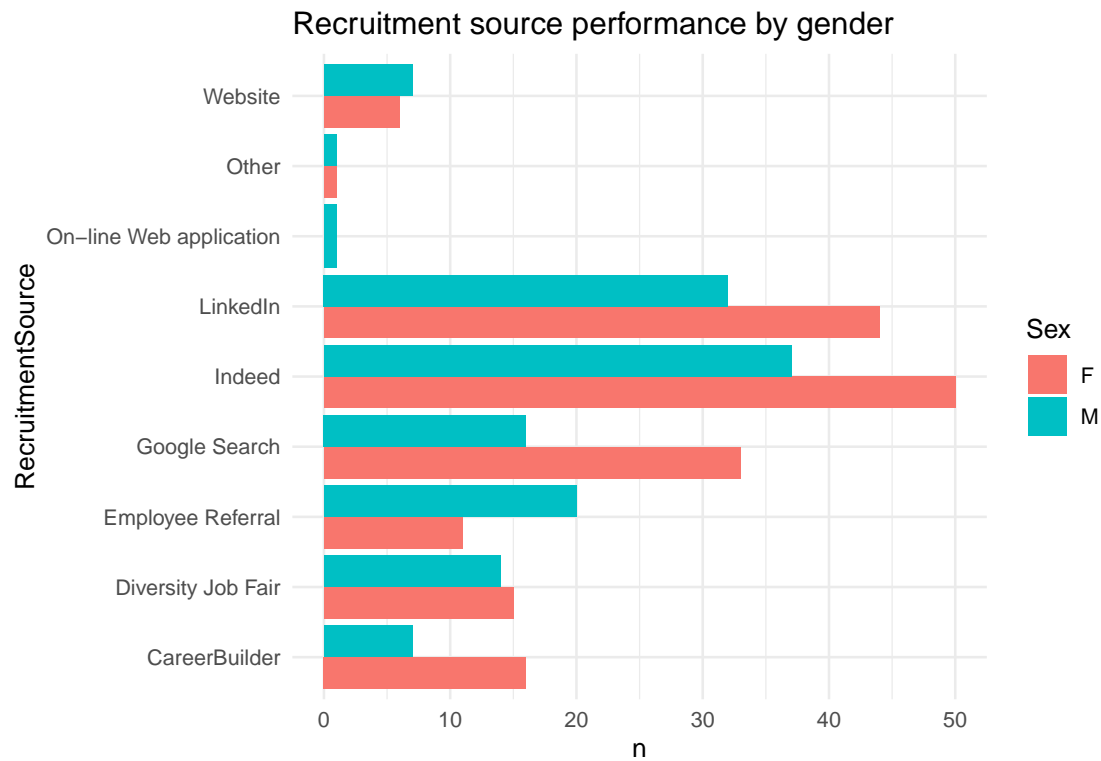
There are some disparity in salary for some roles, some times in favor of women some time the opposite. Over all if we take the whole company in consideration the gap is near to zero.

Average salary by gender and role
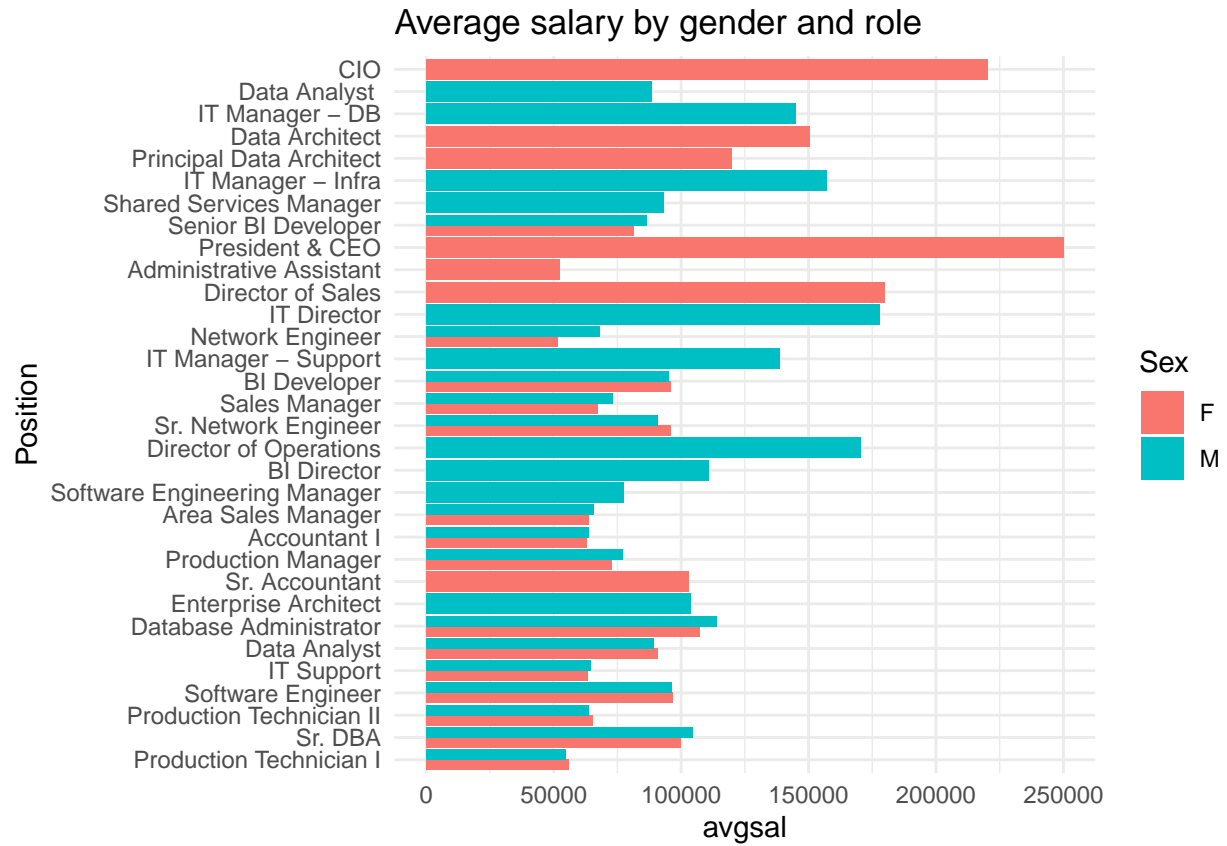
Women has the highest rate of contract termination with the 3 main reasons in other position research, more money and unhappy in the company.
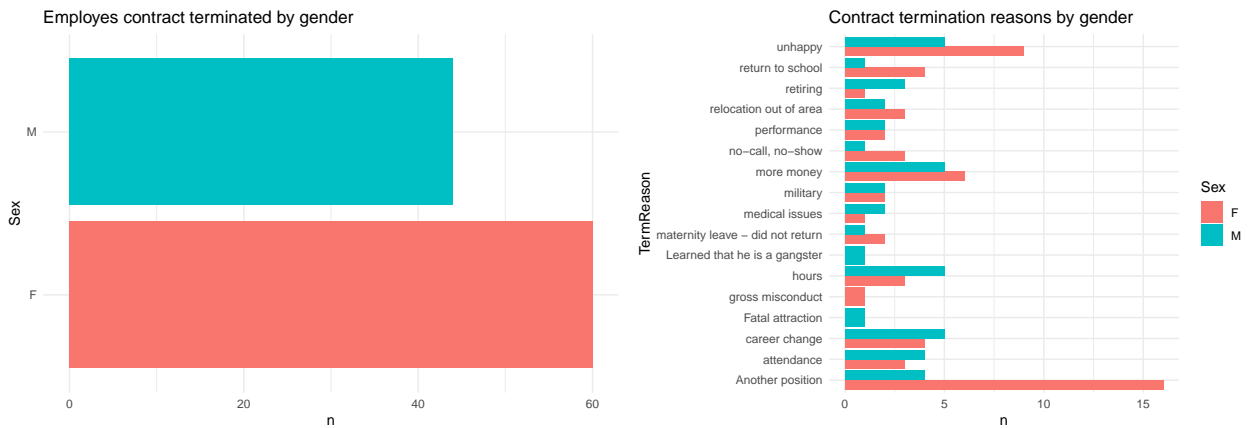
The most effective recruitment source for women and for men was the same (Indeed).

Satisfaction by gender was as follows:

Employes satisfaction by gender

Employes satisfaction level by gender and salary level